



Information Systems

Big Data Analytics

Presented by: Dr Sherin El Gokhy



Module 1 – Introduction to Big Data Analytics



Introduction



Analytics Lifecycle



Basic Methods



Adv. Methods



Tools



Lab

Module 1: Introduction to Big Data Analytics

The Data Scientist

During this part the following topics are covered:

- Key Roles of the New Big Data Ecosystem
- Profile of a Data Scientist

Skills Needed In the New Data Ecosystem



Your Thoughts?

- What new **skill sets** do you need to take advantage of the big data sets in the loan processing improvement case study?
- Do most large organizations have people with these **skill sets**?
- If so, **who are they**?

Three Key Roles of the New Data Ecosystem

Data Scientists

Deep Analytical Talent

People with advanced training in quantitative disciplines, such as mathematics, statistics, and machine learning.

Analysts & Data Savvy Managers

Data Savvy Professionals

People with a basic knowledge of statistics and/or machine learning, who can define key questions that can be answered using advanced analytics

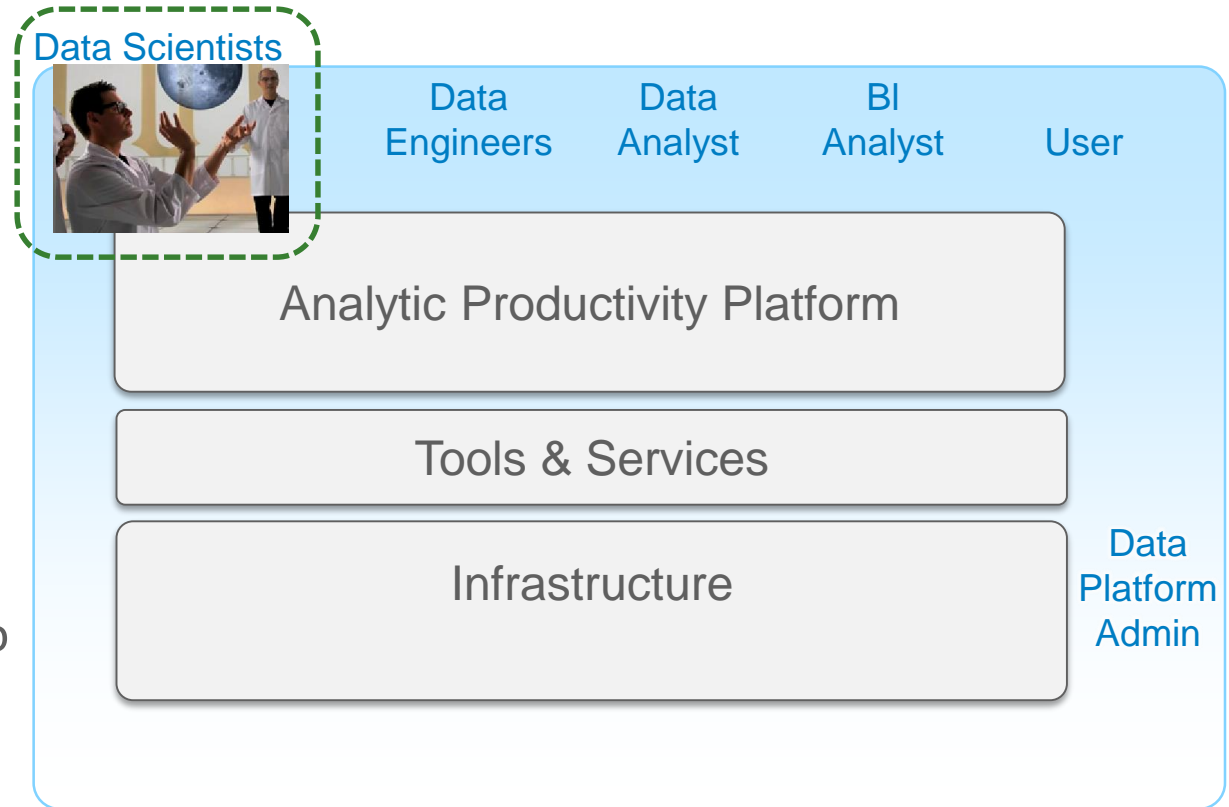
Technology & Data Enablers

People providing technical expertise to support analytical projects. Skills sets including computer programming and database administration

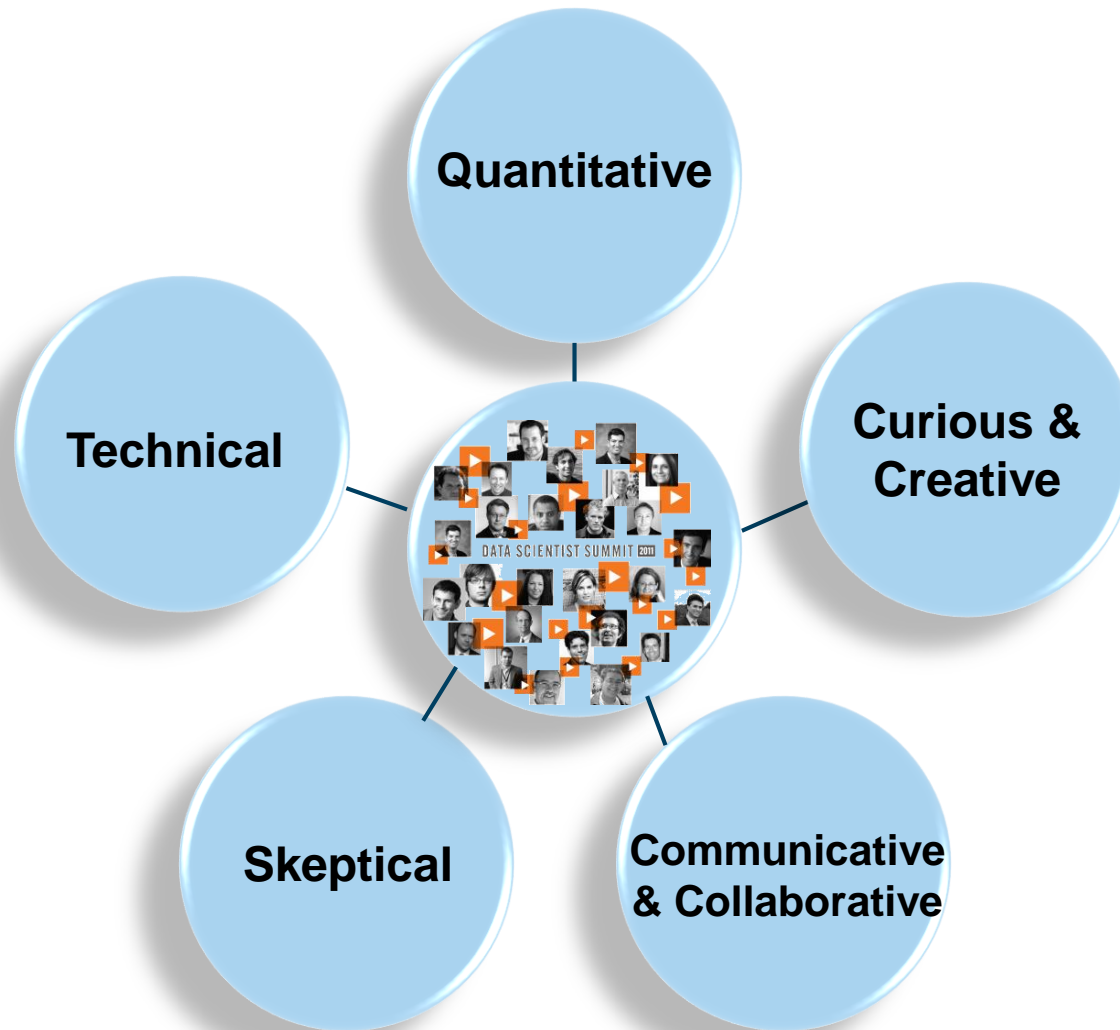
Note: Figures above reflect a projected talent gap in US in 2018, as shown in McKinsey May 2011 article *Big Data: The next frontier for innovation, competition, and productivity*

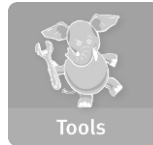
Data Scientist Key Activities

- Reframe business challenges as analytics challenges
- Design, implement and deploy statistical models and data mining techniques on big data
- Create insights that lead to actionable recommendations



Profile of a Data Scientist





Module 1: Introduction to Big Data Analytics

Summary

During this part the following topics were covered:

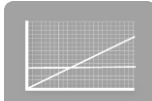
- Key Roles of the New Big Data Ecosystem
- Profile of a Data Scientist



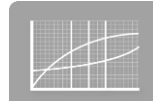
Introduction



Analytics Lifecycle



Basic Methods



Adv. Methods



Tools



Lab

Module 1: Introduction to Big Data Analytics

Big Data Analytics in Industry Verticals

During this part we cover the following representative examples:

- Health Care
- Public Services
- Life Sciences
- IT Infrastructure
- Online Services

Big Data Analytics: Industry Examples

- 1 Health Care
 - Reducing Cost of Care
- 2 Public Services
 - Preventing Pandemics
- 3 Life Sciences
 - Genomic Mapping
- 4 IT Infrastructure
 - Unstructured Data Analysis
- 5 Online Services
 - Social Media for Professionals



1 Big Data Analytics: *Healthcare*



Situation

- Poor police response and problems with medical care, triggered by shooting of a student
- The event drove local doctor to map crime data and examine local health care

Use of Big Data

- Dr. Jeffrey Brenner generated his own crime maps from medical billing records of 3 hospitals

Key Outcomes

- City hospitals provided expensive care, low quality care
- Reduced hospital costs by 56% by realizing that 80% of city's medical costs came from 13% of its residents, mainly low-income or elderly
- Now offers preventative care over the phone or through home visits



Situation

- Threat of global pandemics has increased exponentially
- Pandemics spreads at faster rates, more resistant to antibiotics

Use of Big Data

- Created a network of viral listening posts
- Combines data from viral discovery in the field, research in disease hotspots, and social media trends
- Using Big Data to make accurate predications on spread of new pandemics

Key Outcomes

- Identified a fifth form of human malaria, including its origin
- Identified why efforts failed to control swine flu
- Proposing more proactive approaches to preventing outbreaks



Situation

- Broad Institute (MIT & Harvard) mapping the Human Genome

Use of Big Data

- In 13 yrs, mapped 3 billion genetic base pairs; 8 petabytes
- Developed 30+ software packages, now shared publicly, along with the genomic data

Key Outcomes

- Using genetic mappings to identify cellular mutations causing cancer and other serious diseases
- Innovating how genomic research informs new pharmaceutical drugs



Situation

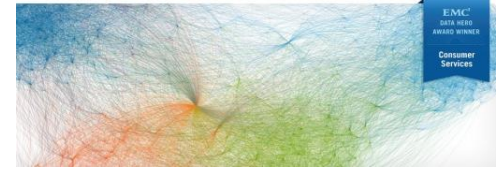
- Explosion of unstructured data required new technology to analyze quickly, and efficiently

Use of Big Data

- Doug Cutting created Hadoop to divide large processing tasks into smaller tasks across many computers
- Analyzes social media data generated by hundreds of thousands of users

Key Outcomes

- New York Times used Hadoop to transform its entire public archive, from 1851 to 1922, into 11 million PDF files in 24 hrs.
- Applications range from social media, sentiment analysis, wartime chatter, natural language processing



Situation

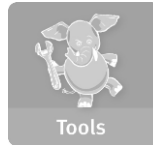
- Opportunity to create social media space for professionals

Use of Big Data

- Collects and analyzes data from over 100 million users
- Adding 1 million new users per week

Key Outcomes

- LinkedIn Skills, InMaps, Job Recommendations, Recruiting
- Established a diverse data scientist group, as founder believes this is the start of Big Data revolution



Module 1: Introduction to Big Data Analytics

Summary

During this part the following representative examples were covered:

- Health Care
- Public Services
- Life Sciences
- IT Infrastructure
- Online Services

Check Your Knowledge



Your Thoughts?

1. What are the 3 characteristics of Big Data, and the main considerations in processing Big Data?
2. What is an analytic sandbox – Data EcoSystem?
3. Explain the difference between Business Intelligence and Data Science.
4. Describe the challenges of the current analytical architecture for Data Scientists.
5. What are the key skill sets and behavioral characteristics of a Data Scientist?



Introduction



Analytics Lifecycle



Basic Methods



Adv. Methods



Tools

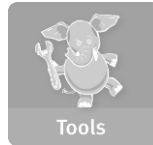


Lab

Module 1: Summary

Key points covered in this module:

- Big data was defined
- Four business drivers for advanced analytics were identified
- The techniques for Business Intelligence were distinguished from those of Data Science
- The role of the Data Scientist within the new big data ecosystem was described
- Multiple illustrative examples of big data opportunities were cited



Thanks